bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1 # Metagenomic sequencing reveals a lack of virus
2 # exchange between native and invasive freshwater fish
3 # across the Murray-Darling Basin, Australia

4

5 Vincenzo A. Costa[1], Jonathon C.O. Mifsud[1], Dean Gilligan[2], Jane E. Williamson[1], Edward C.
6 Holmes[3], Jemma L. Geoghegan[1,4,5]

7

8

9 [1]Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia.

10 [2]NSW Department of Primary Industries, Batemans Bay Fisheries Office, Batemans Bay
11 2536, Australia

12 [3]Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre,
13 School of Life and Environmental Sciences and Sydney Medical School, The University of
14 Sydney, Sydney, NSW 2006, Australia.

15 [4]Department of Microbiology and Immunology, University of Otago, Dunedin 9016, New
16 Zealand.

17 [5]Institute of Environmental Science and Research, Wellington, New Zealand.

18

19 Corresponding author: Jemma Geoghegan; jemma.geoghegan@otago.ac.nz

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

20 # Abstract

21 Biological invasions are among the biggest threats to freshwater biodiversity. This is
22 increasingly relevant in the Murray-Darling Basin, Australia, particularly since the
23 introduction of the common carp (*Cyprinus carpio*). This invasive species now occupies up
24 to 90% of fish biomass, with hugely detrimental impacts on native fauna and flora. To
25 address the ongoing impacts of carp, *cyprinid herpesvirus 3* (*CyHV-3*) has been proposed
26 as a potentially effective biological control. Crucially, however, it is unknown whether *CyHV-*
27 *3* and other cyprinid herpesviruses already exist in the Murray-Darling. Further, little is
28 known about those viruses that naturally occur in wild freshwater fauna, and the frequency
29 with which these viruses jump species boundaries. To document the evolution and diversity
30 of freshwater fish viromes and better understand the ecological context to the proposed
31 introduction of *CyHV-3*, we performed a meta-transcriptomic viral survey of invasive and
32 native fish across the Murray-Darling Basin, covering over 2,200 km of the river system.
33 Across a total of 36 RNA libraries representing 10 species, we failed to detect *CyHV-3* nor
34 any closely related viruses. Rather, meta-transcriptomic analysis identified 18 vertebrate-
35 associated viruses that could be assigned to the *Arenaviridae, Astroviridae, Bornaviridae,*
36 *Caliciviridae, Coronaviridae, Chuviridae, Flaviviridae, Hantaviridae, Hepeviridae,*
37 *Paramyxoviridae, Picornaviridae, Poxviridae, Reoviridae* and *Rhabdoviridae* families, and a
38 further 27 that were deemed to be associated with non-vertebrate hosts. Notably, we
39 revealed a marked lack of viruses that are shared among invasive and native fishes
40 sampled here, suggesting that there is little virus transmission from common carp to native
41 fish species. Overall, this study provides the first data on the viruses naturally circulating in
42 a major river system and supports the notion that fish harbour a large diversity of viruses
43 with often deep evolutionary histories.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

44 # Author Summary

45 The ongoing invasion of the common carp in the Murray-Darling Basin, Australia, has
46 wreaked havoc on native freshwater ecosystems. This has stimulated research into the
47 possible biological control of invasive carp through the deliberate release of the virus
48 *cyprinid herpesvirus 3 (CyHV-3).* However, little is known on the diversity of viruses that
49 naturally circulate in wild freshwater fauna, whether these viruses are transmitted between
50 invasive and native species, nor if *CyHV-3* or other cyprinid herpesviruses are already
51 present in the basin. To address these fundamental questions we employed meta-
52 transcriptomic next-generation sequencing to characterise the total assemblage of viruses
53 (i.e. the viromes) in three invasive and seven native fish species cohabiting at 10 sites
54 across 2,200 km of the river system. From this analysis we identified 18 vertebrate-
55 associated viruses across 14 viral families, yet a marked lack of virus transmission between
56 invasive and native species. Importantly, no *CyHV-3* was detected. This study shows that
57 freshwater fish harbour a high diversity and abundance of viruses, that viruses have likely
58 been associated with fish for millennia, and that there is likely little direct virus transmission
59 between introduced and native species.

60

61 # Introduction

62 Anthropogenic stressors such as pollution, climate change and the introduction of exotic
63 species continue to pose a significant threat to freshwater habitats, with almost one third of
64 all fish species threatened by extinction [1]. The Murray-Darling Basin, the largest
65 freshwater river system in Australia, harbours at least 12 exotic freshwater fish species [2].
66 Key among these are eastern mosquitofish (*Gambusia holbrooki*), redfin perch (*Perca*
67 *fluviatilis*) and, most notably, common carp (*Cyprinus carpio*) [2]. Common carp (also known
68 as European carp) were initially introduced into Australia during the mid-1800s for
69 aquaculture operations and again on several occasions throughout the 1900s [3, 4]. During
70 extensive flooding events during the 1970s, carp spread across much of the basin and now
71 represent up to 90% of total fish biomass in the basin [4].

72 The invasion of carp has been hugely detrimental to Australian freshwater ecosystems [4].
73 Impacts include increased water turbidity, decreased light penetration, erosion of
74 riverbanks, changes in the abundance and diversity of native invertebrate communities [3,
75 4, 5] and outcompeting native fish species for habitat and resources [2]. Several control

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

76 methods have been proposed to control invasive carp; nevertheless, their resilience and
77 high fecundity create significant challenges [6]. This has stimulated research into biological
78 control methods, such as deployment of the virus *cyprinid herpesvirus 3* (*CyHV-3*) [7, 8].

79 *CyHV-3* is a double-stranded DNA virus (family *Alloherpesviridae,* order *Herpesvirales*) first
80 isolated from farmed carp in the late 1990s [9]. Since its discovery, it has been responsible
81 for large disease outbreaks worldwide with a mortality rate of up to 80% in domestic carp
82 [10]. CyHV-3 is transmitted horizontally through direct contact with skin lesions or secretion
83 of viral particles in freshwater where it can survive for up to three days [11]. The host range
84 of CyHV-3 is currently limited to koi and common carp [9]. While *CyHV-3* DNA has been
85 identified in goldfish (*Carrasius auratus*) [12], it is still relatively unclear whether infection
86 occurs in these species [14, 15].

87 Initial laboratory trials suggest that *CyHV-3* is safe for non-target species [7]. However, little
88 is known about the viruses that naturally circulate in Australian native freshwater fauna,
89 including any prior evidence for the existence of *CyHV-3* [13], nor on the time-scales and
90 frequency with which viruses jump between fish hosts. To completely assess the safety and
91 efficacy of any virus biocontrol agent, including *CyHV-3*, a comprehensive assessment of
92 the viruses that naturally infect both native and invasive species is required.

93 Following the advent of meta-transcriptomic sequencing, it is now possible to characterise
94 the entire set of viruses – the virome – within a given host [16, 17]. Fish, in particular,
95 harbour a high abundance and diversity of viruses often with deep evolutionary histories
96 [18, 19]. However, despite the antiquity and diversity of fish viruses, there are few studies of
97 virus diversity and evolution in wild freshwater fish populations, particularly in the context of
98 biological invasions.

99 Determining the viromes of invasive freshwater fish like the common carp will enhance our
100 understanding of the broad-scale factors that influence virus emergence and evolution. As
101 the date and site of their introduction is well-documented in Australian waters, these
102 species can potentially provide important information on the both rate of cross-species
103 transmission and how frequently viruses might move between invasive and native species.
104 In addition, despite representing a small fraction of the earth's surface water, freshwater
105 environments serve as a habitat for 40-50% of total fish species, harbouring the greatest
106 biodiversity per land area [20]. Such habitats are subject to rapid environmental change,
107 which may significantly impact species connectivity [21]. Since contact and exposure
108 between hosts are vital for cross-species transmission of viruses [22], these species may
109 also inform us on the ecological factors that impact virome composition within a given host.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

110 Herein, we performed a meta-transcriptomic viral survey of invasive and native freshwater
111 fish species across the Murray-Darling Basin in Australia to document the diversity and
112 evolution of freshwater fish viromes and, from this, better understand the ecological drivers
113 of virus evolution and emergence. To the best of our knowledge this is the largest survey of
114 freshwater fish viruses undertaken to date. In particular, we aimed to determine whether
115 *CyHV-3* is already present in common carp in Australia [13], and whether there is evidence
116 for transmission of existing viruses between exotic and native species. As such, we provide
117 important information on the ecological and evolutionary context for the potential release of
118 future virus biocontrols.

119

120 Methods

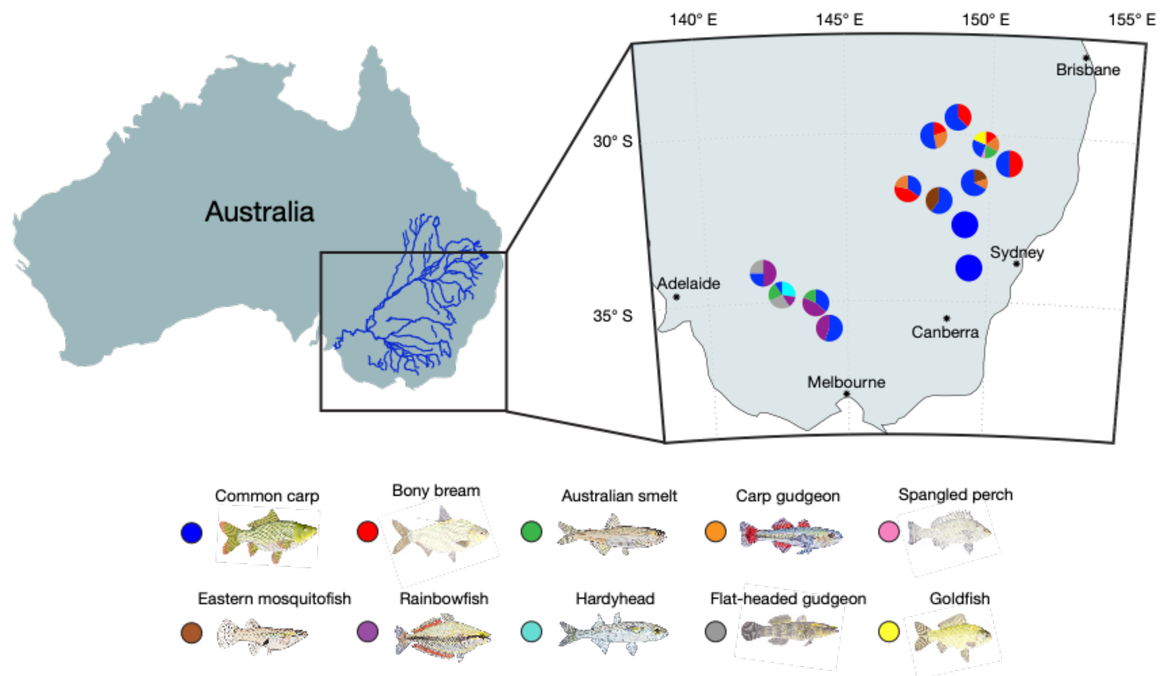
121 Ethics

122 Fish sampling was conducted with animal ethics approval (ref: 2019/035) from the Animal
123 Ethics Committee (AEC) at Macquarie University, Sydney, NSW. Biosafety was approved by
124 Macquarie University (ref: 5201700856).

125 Sample collection

126 We compared the viromes of native and invasive fish species occupying different areas
127 across the Murray-Darling Basin, Australia (Figure 1). Sampling occurred between January
128 and March 2020. A total of seven native fish species were collected: bony herring
129 (*Nematalosa erebi*), spangled perch (*Leiopotherapon unicolor*), Australian smelt (*Retropinna*
130 *semoni*), Murray-Darling rainbowfish (*Melanotaenia fluviatilis*), flat-headed gudgeon
131 (*Philypnodon grandiceps*), western carp-gudgeon (*Hypseleotris spp.*) and unspecked
132 hardyhead (*Craterocephalus fulvus*). Three species of invasive fish were also collected:
133 common carp (*Cyprinus carpio*), goldfish (*Carassius auratus*) and eastern mosquitofish
134 (*Gambusia holbrooki)*. All fish caught were apparently healthy, with no signs of disease. Fish
135 were caught using boat electrofishing, euthanized and dissected immediately upon capture.
136 Tissue specimens (liver and gills) were placed in RNALater and stored in a portable -80°C
137 freezer, then later in a -80°C freezer in the laboratory at Macquarie University, Sydney.
138 Tissue selection was based on previous studies [18, 37, 44], which show that liver and gill
139 tissue serve as a rich source of viruses. To facilitate virus discovery, multiple individuals (1-
140 10) were pooled according to species and the location in which they were captured (SI
141 Table 2; Figure 1).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



142

143 **Figure 1.** Fish sampling locations. Map illustrating freshwater sampling locations across the
144 Murray-Darling Basin, Australia. Pie-charts show the abundance and diversity of fish
145 species captured at each site. Colours correspond to the fish species sampled as
146 illustrated below the map.

147

148 Total RNA extraction and transcriptome sequencing

149 Frozen samples of liver and gill tissue were placed together in 600μl of lysis buffer
150 containing 0.5% foaming reagent (Reagent DX, Qiagen) and 1% of ß-mercaptoethanol
151 (Sigma-Aldrich). Submerged tissue samples were homogenized with TissueRuptor (Qiagen)
152 for one minute at 5000 rpm. To further homogenise tissue samples and remove tissue
153 residues, the homogenate was centrifuged at full speed for three minutes. The homogenate
154 was carefully removed and RNA from the clear supernatant was extracted using the
155 RNeasy Plus Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol.
156 Extracted RNA was quantified using NanoDrop (ThermoFisher) and RNA from each species
157 was pooled corresponding to the site in which they were captured, resulting in 36 sample
158 libraries (SI Table 2). RNA libraries were constructed using the Truseq Total RNA Library
159 Preparation Protocol (Illumina). To enhance viral discovery and reduce the presence of non-
160 viral reads, host ribosomal RNA (rRNA) was depleted using the Ribo-Zero-Gold Kit (Illumina)
161 and paired-end sequencing (150 bp) was performed on the NovaSeq 500 platform

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

162 (Illlumina). Sample library construction, rRNA depletion and RNA sequencing were
163 performed at the Australian Genome Research Facility (AGRF).

164 **Virus discovery**

165 Raw Illumina sequence reads (forward and reverse) were initially quality trimmed with
166 Trimmomatic v.0.39 [23] then assembled into contigs *de novo* using Trinity RNA-seq v.2.8.5
167 [24], with the default parameter settings. Assembled contigs were annotated and compared
168 against the NCBI nucleotide (nt) and non-redundant protein (nr) databases with an e-value
169 threshold of $1x10^{-5}$ using BLASTn and Diamond (BLASTX) [25]. To initially distinguish
170 between invertebrate and vertebrate-associated viruses, contigs that matched viral
171 sequences were inspected using Geneious v.11.1.5 [26] and translated into amino acid
172 sequences. Amino acid sequences were then used as a single query in additional sequence
173 comparisons against the NCBI nt and nr databases using BLAST algorithms. This method
174 was also used to remove false positives (e.g. host genes and endogenous viral elements)
175 from our analyses. To help exclude instances of index hopping, viral sequences that were
176 identified in multiple libraries were also inspected using Geneious Prime
177 (www.geneious.com) and amino acid pairwise alignments between viral sequences were
178 performed with Multiple Alignment using Fast Fourier Transform (MAFFT) v.7.450 [27], using
179 the E-INS-i algorithm. Abundances of identical viral transcripts were then calculated (see
180 below) and sequences that were present at frequency of <1% of that of the number of
181 reads present in the dominant library were excluded. To determine whether a virus was
182 novel, we followed the broad criteria specified by The International Committee on
183 Taxonomy of Viruses (ICTV) (http://www.ictvonline.org/).

184 **Inferring the evolutionary history of novel viral sequences**

185 To determine the evolutionary history of the viruses identified in this study and further
186 distinguish between vertebrate and invertebrate-associated viruses (which are usually
187 phylogenetically distinct), we estimated phylogenetic trees using amino acid sequences of
188 stable genomic regions such as RNA-dependant RNA polymerase (RdRp) or DNA
189 polymerase for DNA viruses. To this end, we combined our sequences with background
190 sequences for each respective virus family taken from NCBI/GenBank. Amino acid
191 sequences were aligned with MAFFT v.7.450 [27] using the E-INS-i algorithm. To remove
192 ambiguous regions in the sequence alignment, amino acid sequences were trimmed using
193 trimAl v.1.2 [33]. To estimate phylogenetic trees, selection of the best-fit model of amino
194 acid substitution was determined using the Akaike information criterion (AIC), corrected
195 Akaike information criterion (AICc), and the Bayesian information criterion (BIC) with the

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

196 ModelFinder function (-m MFP) in IQ-TREE [34, 35]. Sequence data were analysed using a
197 maximum likelihood (ML) approach in IQ-TREE, with 1000 bootstrap replicates.
198 Phylogenetic trees were annotated with FigTree v.1.4.2. and further edited using Adobe
199 Illustrator (https://www.adobe.com).

200 **Virome composition**

201 To quantify the relative abundance of viral transcripts within the host transcriptome, the
202 RNA-Seq by Expectation (RSEM) value was estimated using Trinity [24], and raw counts
203 from each transcript were standardised against the total number of reads within the given
204 sequencing library. We also used this approach to estimate the relative abundance of a
205 host reference gene, ribosomal protein S13 (RPS13), that is stably expressed in fish. To
206 assess any differences in virome composition between hosts and sites, we calculated alpha
207 diversity (virome richness and Shannon diversity) using Rhea packages [28]. Generalised
208 linear models (GLM) were used to identify the impact of host taxonomy (i.e. species), host
209 geography (i.e. site), water temperature, water pH, water turbidity and species origin (i.e.
210 invasive or native) on both vertebrate-associated virus composition (abundance, richness
211 and diversity) as well as those viruses likely associated with non-fish hosts: the latter should
212 not be affected by aspects of fish biology and hence effectively constitute a negative
213 control. All GLM models were tested using a likelihood-ratio test ($\chi 2$) and a Tukey's post
214 hoc analysis (glht) was performed using the *multcomp* package [29]. To assess viral
215 diversity between samples, we calculated beta diversity using a Bray Curtis dissimilarity
216 with the *phyloseq* package [30]. Differences in virome composition between native and
217 invasive species were calculated using permanova (Adonis test), with the *vegan* package
218 [31]. All statistical analyses were carried out on RStudio V1.2.1335 and plotted using the
219 *ggplot2* package [32].

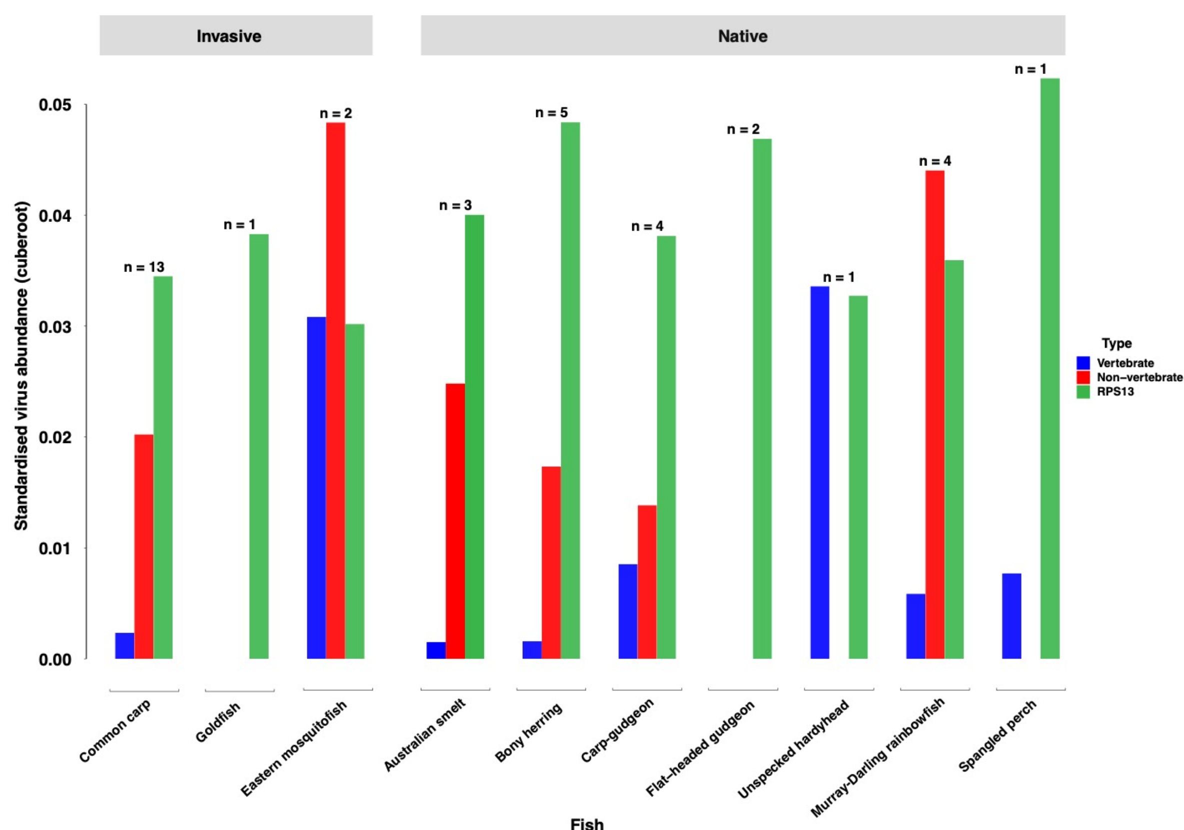
220

221 # Results

222 We characterised the viromes of ten freshwater ray-finned fish species across seven
223 taxonomic orders (two invasive and five native) at 13 locations across the Murray-Darling
224 Basin in Australia. Total RNA-sequencing was performed on 36 libraries, resulting in a
225 median of 76,528,534 (range 66,015,138 – 95,168,951) reads per library. *De novo* assembly
226 of the sequencing reads resulted in a median of 617,588 contigs (range 198,446 –
227 1,989,596) per library, with a total of 23,976,218 contigs generated. Analysis of the host

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

228 reference gene, RPS13, revealed abundances of 0.000001 – 0.0002%, suggesting an
229 inconsistent sequence coverage across all RNA libraries (Figure 2).



230

231 **Figure 2**. Mean standardised viral abundance across all libraries. Clustered bar chart
232 reveals differences in viral abundance between invasive and native fish species. Blue bars
233 represent vertebrate-associated viruses; red bars represent non-vertebrate associated
234 viruses; and green bars represent host reference gene RPS13. Number of sequencing
235 libraries for each fish species is displayed above bars.

236

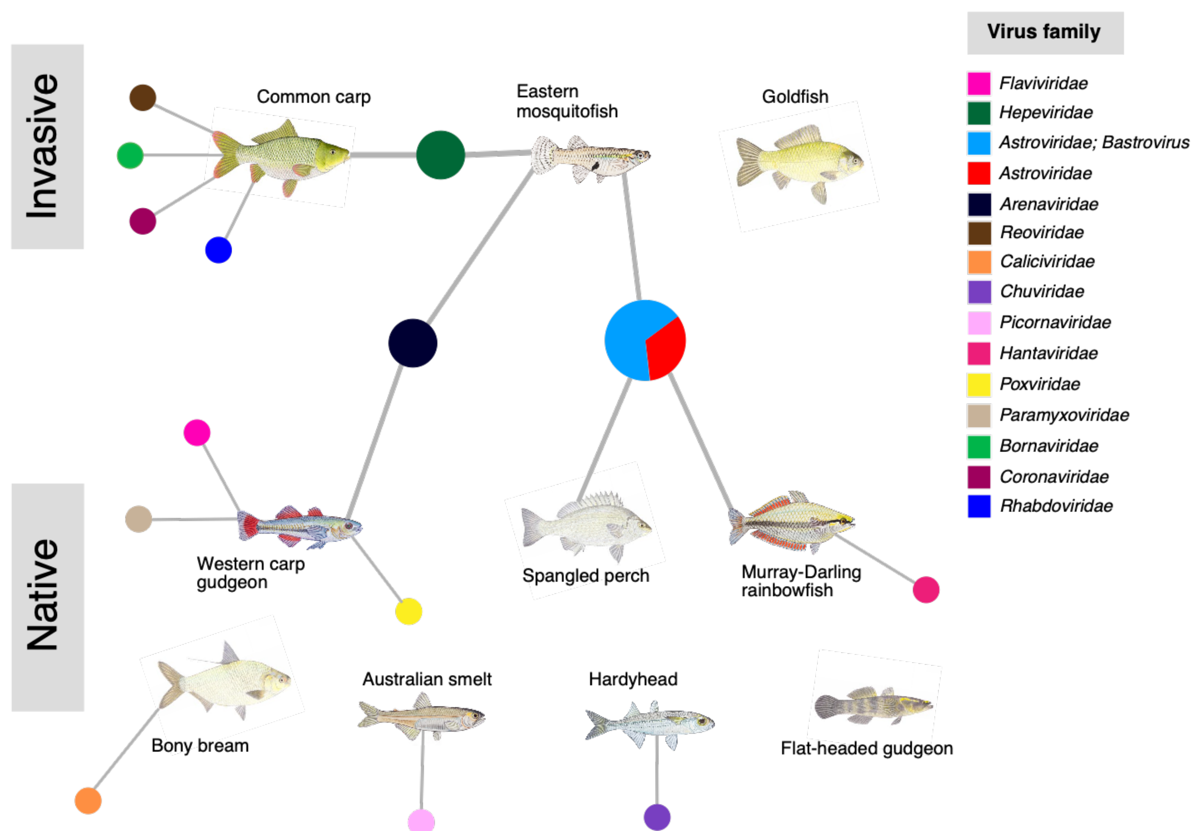
237 **Abundance and diversity of viruses**

238 We identified 18 viral sequences that were associated with vertebrate hosts and a further
239 27 that were likely associated with algae, invertebrates and protists in the freshwater
240 environment (SI Figures 1 and 2). Because such non-vertebrate viruses were likely derived
241 from diet or contamination of gill tissue, we primarily focused on vertebrate-associated
242 viruses.

243 Among the likely vertebrate-associated viruses, we identified viral sequences from 14 viral
244 families. With the exception of a novel poxvirus (family *Poxviridae*), a double-stranded DNA
245 virus, all the viruses identified possessed RNA genomes. The most abundant vertebrate-

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

246 associated viral transcripts were those assigned to the *Arenaviridae* (49% of all vertebrate-
247 associated viruses), *Hepeviridae* (20%), *Chuviridae* (21%), *Astroviridae* (3%) and *Flaviviridae*
248 (2%) families. Other likely vertebrate viral transcripts detected were assigned to the
249 *Coronaviridae* (<1%) *Caliciviridae* (<1%), *Picornaviridae* (<1%), *Paramyxoviridae* (<1%),
250 *Hantaviridae* (<1%), *Bornaviridae* (<1%), *Poxviridae* (<1%), *Reoviridae* (<1%) and
251 *Rhabdoviridae* (<1%) families. The most common vertebrate-associated viruses identified
252 were astroviruses, detected in three host species (eastern mosquitofish, Murray-Darling
253 rainbowfish, spangled perch). In addition, arenaviruses were detected in two host species
254 (western carp-gudgeon, eastern mosquitofish) along with hepeviruses (common carp,
255 eastern mosquitofish). All other viruses were identified in one host species (Figure 3).



256

257 **Figure 3** Network diagram displaying vertebrate-associated viruses identified in native and
258 invasive freshwater fish. Colours of each node represents a virus family. Both goldfish and
259 flat-headed gudgeon contained no vertebrate-associated viruses.

260

261 Among the viruses likely associated with non-vertebrate hosts (i.e. those infecting
262 arthropods, fungi, plants and protozoans), a large proportion (70%) were unclassified,
263 comprising picorna-like viruses, rhabdo-like viruses, tombus-like viruses and narna-like

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

264 viruses [45] (SI Figure 1). We also detected viral transcripts that could be assigned to the
265 *Nodaviridae* (27.1%), *Permutotetraviridae* (1.3%), *Dicistroviridae* (1.2%) and *Phenuiviridae*
266 (<1%) families. Although viruses within the *Nodaviridae* have been shown to infect fish [46],
267 all of the nodavirus sequences identified here clustered with viruses from invertebrate hosts
268 (SI Figure 2), strongly suggesting they were similarly associated with fish diet or
269 contamination of gill tissue.

270 **Phylogenetic relationships of vertebrate-associated viruses**

271 To infer the phylogenetic relationships and hence the evolutionary history of the viruses
272 newly identified here, we focused on stable genomic regions such as the RdRp in RNA
273 viruses and DNA polymerase in the case of the novel poxvirus. Using these genomic
274 regions, we identified seven negative-sense single-stranded RNA (-ssRNA) viruses (families
275 *Arenaviridae, Bornaviridae, Chuviridae, Hantaviridae, Paramyxoviridae, Rhabdoviridae*), nine
276 positive-sense single-stranded RNA (+ssRNA) viruses (families *Astroviridae, Caliciviridae,*
277 *Coronaviridae, Flaviviridae, Hepeviridae, Picornaviridae*), one double-stranded RNA (dsRNA)
278 virus (family *Reoviridae*) and one double-stranded DNA (dsDNA) virus (family *Poxviridae*).
279 We now describe each of these groups in turn.

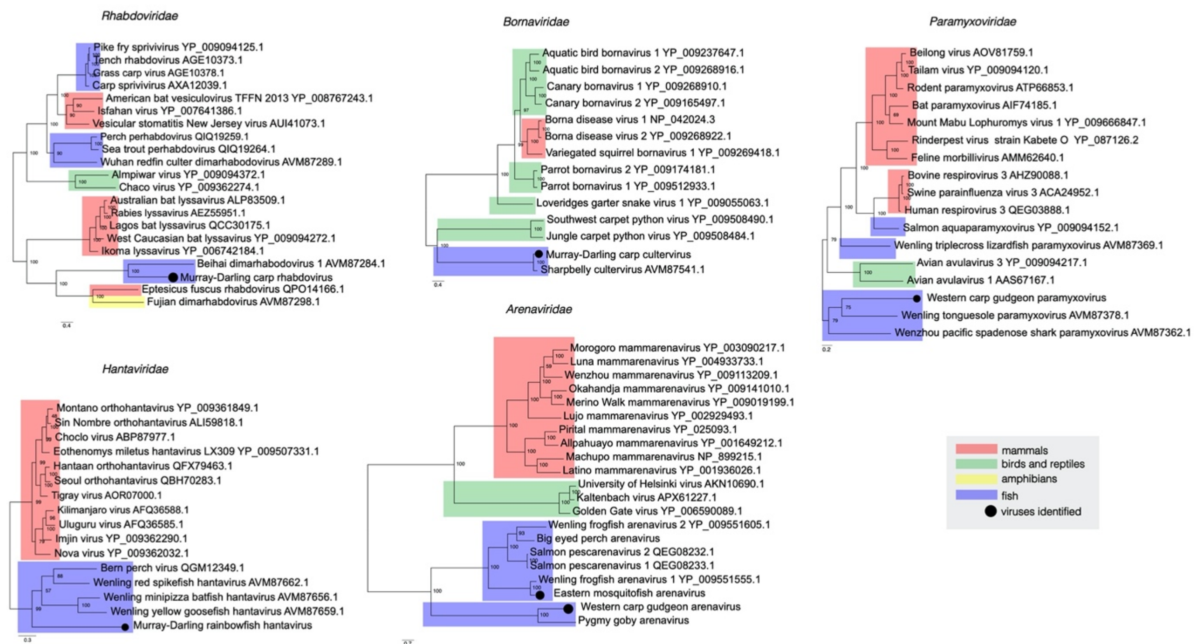
280 ***Negative-sense single-stranded (-ssRNA) viruses***

281 We identified -ssRNA viruses that occupied phylogenetic positions that were broadly
282 indicative of long-term virus-host co-divergence, with many fish viruses falling basal to
283 reptile and mammalian viruses (Figure 4). Notably, we identified two novel arenaviruses that
284 clustered with members of the newly formed *Antennavirus* genus that includes fish hosts
285 [18, 36]. *Western carp-gudgeon arenavirus*, found at Narrabri creek shared 36.9% amino
286 acid sequence similarity with its closest relative, *Wenling frogfish arenavirus 1* [18] and
287 grouped with recently discovered arenaviruses from pygmy gobies (*Eviota zebrina*) sampled
288 from an Australian coral reef [37]. *Eastern mosquitofish arenavirus* found in the Macquarie
289 River shared 84.5% amino acid sequence similarity with its closest available relative,
290 *Wenling frogfish arenavirus 1* [18].

291 The divergent hantavirus detected in the Murray-Darling rainbowfish falls basal to
292 mammalian hantaviruses (orthohantaviruses) and clustered with members of the
293 *Actantavirinae* and *Agantavirinae* subfamilies that include ray-finned and jawless fish hosts
294 [18, 36] (Figure 4). This virus had only 27.3% amino acid similarity with its closest relative,
295 *Bern perch virus* (NCBI/GenBank: QGM12349.1). Broad patterns of virus-host co-
296 divergence can similarly be seen in the cultervirus identified in carp from Lake Burrendong.
297 BLAST analysis identified *sharpbelly cultervirus* [18] as the closest relative of all genomic

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

298 regions, including the L gene (93% amino acid similarity), glycoprotein (86.3%) and
299 nucleoprotein (92.9%).



300

301 **Figure 4.** Phylogenetic relationships of negative-sense single-stranded vertebrate-
302 associated viruses identified in this study. Viruses identified here are shown as a black
303 circle. Maximum likelihood trees were estimated using amino acid sequences of the RdRp
304 gene and mid-point rooted for clarity only. Bootstrap values are represented as a
305 percentage with branches scaled to amino acid substitutions per site. Tree branches are
306 highlighted to represent host class: red, mammals; green, birds and reptiles; yellow,
307 amphibians; and blue, fish. Tip labels represent virus name and NCBI/GenBank accession
308 numbers.

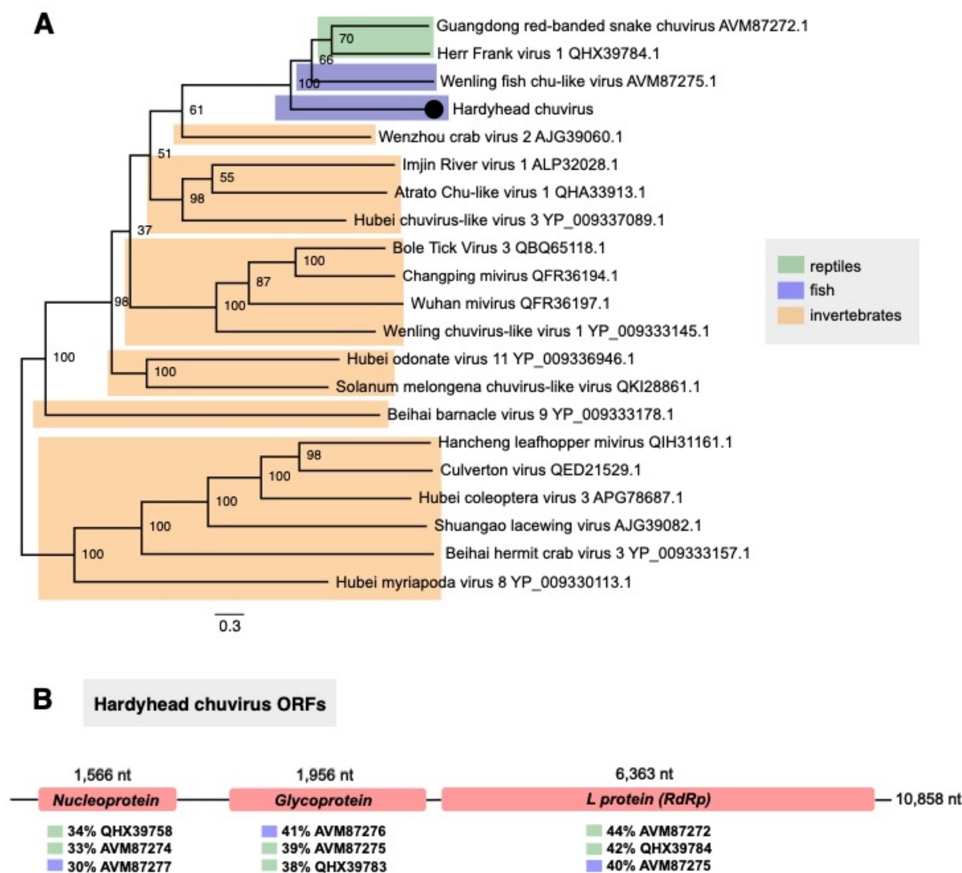
309

310 Our virological survey revealed the complete genome of a novel chuvirus in the unspecked
311 hardyhead in the Edward River. This *Hardyhead chuvirus* displayed three open reading
312 frames (ORFs), representing the L protein (RdRp), glycoprotein and nucleoprotein. Our
313 analysis identified *Guangdong red-banded snake chuvirus* [18] as the closest relative of the
314 L protein (44% amino acid similarity), *Wenling fish chu-like virus* [18] as the closest relative
315 of the glycoprotein (41%), and *Herr Frank virus 1* [41] as the closest relative of the
316 nucleoprotein (34%). *Hardyhead chuvirus* formed a distinct phylogenetic clade with all other
317 vertebrate-associated chuviruses (Figure 5).

318 We also detected a novel paramyxovirus in western carp-gudgeon in the Bogan River. This
319 divergent viral sequence shared 35.2% L gene amino acid similarity with its closest relative,

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

320 *Wenling tonguesole paramyxovirus* (genus *Cynoglossusvirus*, family *Paramyxoviridae*) [18].
321 These viruses grouped with *Wenzhou pacific spadenose shark paramyxovirus* (genus
322 *Scoliodonvirus*), together falling basal to other members of the *Paramyxoviridae* family. In
323 addition, a novel rhabdovirus in common carp similarly formed a distinct clade, basal to
324 other fish-infecting rhabdoviruses. This virus shared 35.7% amino acid L gene sequence
325 similarity with *Beihai dimarhabdovirus* that was also identified in fish [18] and clustered with
326 other dimarhabdoviruses, including those found in the spotted paddle-tail newt from China
327 [18] and the big brown bat (*Eptesicus fuscus*) from the USA (NCBI/Genbank: QPO14166.1).
328 Across all genera within the *Rhabdoviridae*, lyssaviruses were the closest relatives to this
329 clade (Figure 4), with *Murray-Darling carp rhabdovirus* sharing 31.6% amino acid L gene
330 similarity with *rabies lyssavirus* [43].



331

332 **Figure 5.** Phylogenetic relationships and genome organisation of *hardyhead chuvirus*. (A)
333 Phylogenetic relationships between viruses within the *Chuviridae*. Novel chuvirus identified
334 in the unspecked hardyhead is represented as a black circle. The phylogenetic tree was
335 estimated using amino acid sequences of the RdRp gene and midpoint rooted for clarity
336 only. The scale bar represents amino acid substitutions per site. Bootstrap values are

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

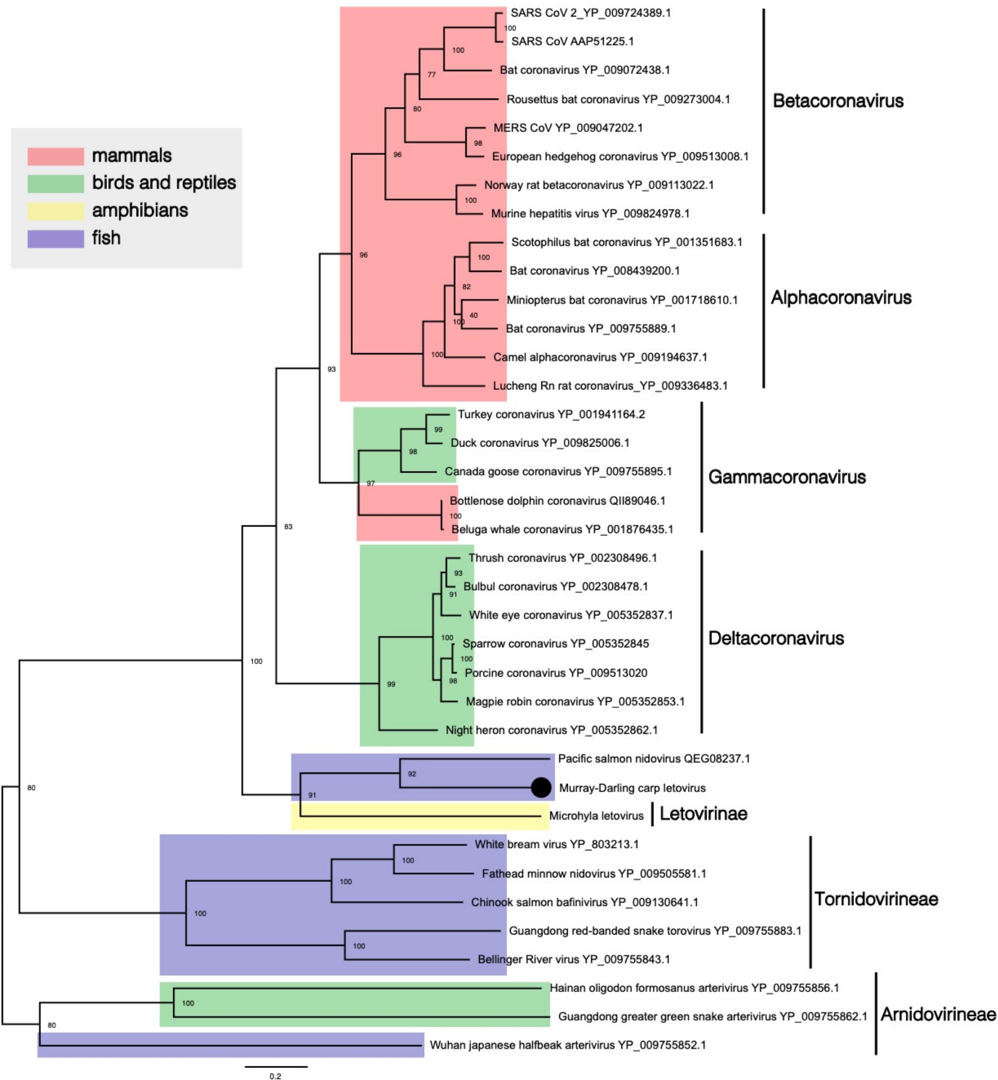
337 shown as a percentage. Tree branches are highlighted to distinguish between vertebrate
338 and invertebrate-associated viruses: green, reptiles; blue, fish; and orange, invertebrates.
339 Tip labels represent virus name and NCBI/GenBank accession numbers. (B) Genome
340 organisation of *hardyhead chuvirus*. Diagram illustrates the structure and length of each
341 genomic segment. Percentages below each segment reveal the closest relatives
342 (NCBI/GenBank accession number), with the L protein more related to reptile chuviruses;
343 glycoprotein more related to fish chuviruses; and nucleoprotein more related to reptile
344 chuviruses.

345

346 ***Positive-sense single-stranded (+ssRNA) viruses***

347 We identified a viral sequence in common carp that shared 50.7% RdRp sequence
348 similarity with *Pacific salmon nidovirus* (family *Coronaviridae*) [60]. *Murray-Darling carp*
349 *letovirus* also exhibited sequence similarity (46.2%) with gammacoronaviruses, including
350 *bottlenose dolphin coronavirus* and *beluga whale coronavirus* [58, 59]. This virus grouped
351 with both *Pacific salmon nidovirus* and *Microhyla letovirus* [81], which together form an
352 outgroup to all other coronaviruses (Figure 6).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



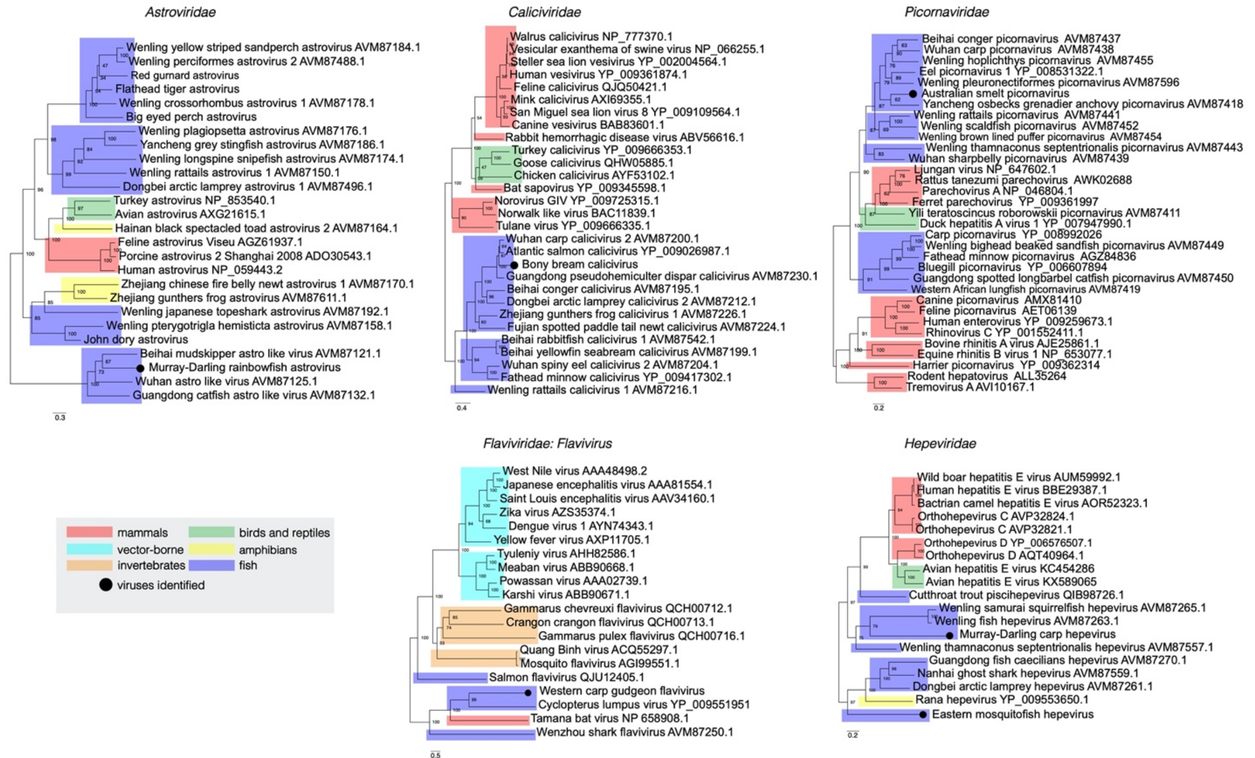
353

354 **Figure 6.** Phylogenetic relationships within the virus order *Nidovirales*. The novel letovirus
355 detected in common carp is labelled with a black circle. The phylogenetic tree was
356 estimated using the orf1ab polypeptide (polymerase associated) and midpoint rooted for
357 clarity alone. Scale bar represents amino acid substitutions per site. Bootstrap values are
358 shown as a percentage. Taxon labels are highlighted: blue, fish; red, mammals; green, birds
359 and reptiles; yellow, amphibians. Virus taxonomic names are labelled to the right. Tip labels
360 represent virus name and NCBI/GenBank IDs.

361

362 We also identified a novel flavivirus (genus *Flavivirus*, family *Flaviviridae*) in western carp-
363 gudgeon in the Bogan River. This viral sequence exhibited 33-36% NS5 amino acid
364 sequence similarity with its closest relatives, *Cyclopterus lumpus virus* [38], *Tamana bat*
365 *virus* [39], *salmon flavivirus* [40] and *Wenzhou shark flavivirus* [18]. All these viruses fall basal
366 to vector-borne viruses within the genus *Flavivirus* (Figure 7).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



367

368 **Figure 7.** Phylogenetic relationships of positive-sense singe-stranded vertebrate-
369 associated viruses identified in this study. Viruses identified here are shown as a black
370 circle. Maximum likelihood trees were estimated using amino acid sequences of the RdRp
371 gene and NS5 gene for the novel flavivirus. Trees were mid-point rooted for clarity only and
372 bootstrap values are represented as a percentage with branches scaled to amino acid
373 substitutions per site. Tree branches are highlighted to represent host class: red, mammals;
374 cyan, vector-borne viruses; green, birds and reptiles; yellow, amphibians; blue, fish; and
375 orange, invertebrates. Tip labels represent virus name and NCBI/GenBank accession IDs.

376

377 Among other positive-sense RNA viruses identified, a novel astrovirus, calicivirus and
378 picornavirus all grouped with other fish hosts and expanded the phylogenetic diversity of
379 these virus families (Figure 7). The novel astrovirus identified in Murray-Darling rainbowfish
380 shared 40% RdRp amino acid similarity with *Wuhan astro-like virus* [18]. This virus
381 clustered with other astro-like viruses discovered in fish, including *Beihai mudskipper astro-*
382 *like virus* [18] and *Guangdong catfish astro-like virus* [18]. This was similarly observed in
383 *Australian smelt picornavirus*, which clustered with picornaviruses found in other freshwater
384 fish, including those from eels (*Anguilla anguilla*) [48] and carp sampled from China [18]. The
385 *Caliciviridae* includes two genera that infect fishes: saloviruses associated with salmonid
386 hosts and minoviruses associated with cyprinid hosts [49]. Recently, several caliciviruses

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

387 have been discovered in ray-finned and jawless fish [18, 50]. The novel calicivirus identified
388 in bony herring expands the diversity of fish viruses as it shared 80% amino acid similarity
389 with *Atlantic salmon calicivirus* [50] and clustered with other freshwater fish caliciviruses,
390 including *Wuhan carp calicivirus* [18] and *Guangdong pseudohemiculter dispar calicivirus*
391 [18].

392 ***Double-stranded RNA (dsRNA) viruses***

393 We identified a novel dsRNA virus in carp in the Castlereagh River that could be assigned
394 to the *Reoviridae*. This divergent virus shared 40% RdRp amino acid similarity with its
395 closest relative, *Wenling scaldfish reovirus* [18], together forming a clade basal to the genus
396 *Aquareovirus* that are known to cause considerable disease in some fish species [42]
397 (Figure 8).

398

399 **Figure 8.** Phylogenetic relationships within the *Reoviridae*. Novel reovirus identified here
400 shown as a black circle. Maximum likelihood tree was estimated using amino acid
401 sequences of the RdRp gene. Tree was mid-point rooted for clarity only and bootstrap
402 values are represented as a percentage with branches scaled to amino acid substitutions
403 per site. Tree branches are highlighted to represent host class: red, mammals; cyan,
404 vector-borne viruses; green, birds and reptiles; blue, fish; and orange, invertebrates. Tip
405 labels represent virus name and NCBI/GenBank accession IDs.

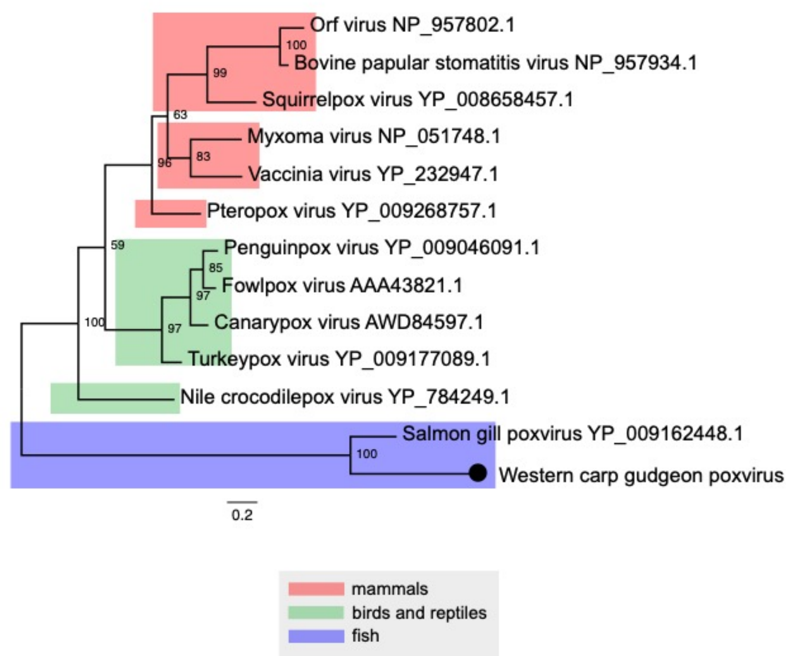
406

407

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

408 ***Double-stranded DNA (dsDNA) viruses***

409 A key observation of our study was the absence of Cyprinid herpesviruses, including *CyHV-*
410 *3*, as well as other *Alloherpesviridae*, in any of the 36 RNA libraries. Similarly, although
411 members of the *Hepadnaviridae* are commonly detected in fish [37, 44, 77, 78] they were
412 notably absent in our samples. The only DNA virus detected in this study was a novel
413 poxvirus (*Poxviridae*) identified in western carp-gudgeon. This virus shared DNA
414 polymerase amino acid sequence similarity with *salmon gill poxvirus* (*SGPV*) (61%) [51]. We
415 also detected other genomic regions such as DNA-dependant RNA polymerase subunit
416 rpo22 (49.5%), DNA-dependant RNA polymerase subunit rpo19 (40%), DNA binding virion
417 core protein I1L (28.1%), A16L (32.9%) and SGPV079 (40%) (SI Table 1). Both *SGPV* and
418 *western carp-gudgeon poxvirus* form a highly divergent clade within the subfamily
419 *Chordopoxvirinae* that is strongly indicative of virus-host co-divergence (Figure 9).



420

421 **Figure 9.** Phylogenetic relationships within the subfamily *Chordopoxvirinae* (family
422 *Poxviridae*). Phylogenetic tree reveals virus-host co-divergence, with fish viruses falling
423 basal to reptilian, avian and mammalian poxviruses. Novel poxvirus identified here shown
424 as a black circle. Maximum likelihood tree was estimated using amino acid sequences of
425 the DNA polymerase gene. Tree was mid-point rooted for clarity only and bootstrap values
426 are represented as a percentage with branches scaled to amino acid substitutions per site.
427 Tree branches are highlighted to represent host class: red, mammals; green, birds and
428 reptiles; blue, fish. Tip labels represent virus name and NCBI/GenBank accession IDs.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

429 **Virome composition, ecological and environmental factors**

430 We next examined whether and how vertebrate virome composition in a range of native and
431 introduced Murray-Darling Basin fish was associated host ecological factors, namely host
432 species, geography (i.e. sampling site), water quality (temperature, pH and turbidity). GLMs
433 revealed host species ($\chi 2 = 7.5^{-6}$, df= 9, p = 0.001) as the best predictor of viral abundance
434 (i.e. the standardised number of viral sequencing reads) (Figure 2). In particular, the eastern
435 mosquitofish had significantly higher viral abundance compared to Australian smelt (Tukey:
436 z=3.976, p=0.002), bony herring (Tukey: z=4.334, p=0.001), western carp-gudgeon (Tukey:
437 z=4.019, p=0.002), common carp (Tukey: z=4.665, p=0.001), flat-headed gudgeon (Tukey:
438 z=3.632, p=0.001), goldfish (Tukey: z=3.632, p=0.001) and rainbowfish (Tukey: z=4.110,
439 p=0.001). However, the high viral abundance in the eastern mosquitofish was driven by one
440 sample containing an extremely high abundance of arenaviruses, accounting for 76% of its
441 total vertebrate virome and 49% of all vertebrate-associated viral reads. We found no
442 evidence for an association between viral abundance and host geography (p = 0.111),
443 water turbidity (p = 0.804), water temperature (p = 0.709) nor water pH (p = 0.141).

444 We calculated alpha diversity to assess any differences in virome composition (abundance
445 and diversity) between hosts and sites. This included the observed virus species richness
446 (the number of viruses found in each sequencing library) and Shannon diversity (both the
447 number of viral families and abundance of viral reads in a given host). We found no
448 association between observed viral species richness and host species (p = 0.286), host
449 geography (p = 0.748), water turbidity (p = 0.826), water temperature (p = 0.625) and water
450 pH (p = 0.115). Similarly, there was also no observed association between Shannon
451 diversity and host taxonomy (p = 0.117), host geography (p = 0.55), water turbidity (p =
452 0.546), water temperature (p = 0.206) and water pH (p = 0.039).

453 **Virome composition of native versus invasive fish species**

454 While carp and native fish species were sampled together at 10 out of 13 sites (Figure 1),
455 they shared no vertebrate-associated viruses. Although we identified hepeviruses in
456 common carp and eastern mosquitofish, these were highly divergent and exhibited only
457 20% amino acid similarity such that they reflect ancient divergence events. These viruses
458 were also distinct in that *Murray-Darling carp hepevirus* clustered with ray-finned fish hosts
459 [18], while *eastern mosquitofish hepevirus* formed a distinct basal clade with amphibian
460 [54], jawless fish and cartilaginous fish hosts [18] (Figure 7).

461 While the native and invasive fish species largely had distinct viromes, two vertebrate-
462 associated virus families were present in both: arenaviruses (western carp-gudgeon and

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

463 eastern mosquitofish) and bastroviruses (spangled perch and eastern mosquitofish) (Figure
464 3). However, the arenaviruses identified in the eastern mosquitofish and western carp-
465 gudgeon were highly divergent, exhibiting only 27.9% amino acid similarity, with *western*
466 *carp-gudgeon arenavirus* falling basal to *eastern mosquitofish arenavirus* (Figure 4). The
467 bastroviruses detected in eastern mosquitofish and spangled perch shared 57.1% RdRp
468 amino acid similarity and formed a distinct clade with other bastrovirus sequences
469 identified in *Culex* mosquitos [52], bats (NCBI/GenBank: NC_035471.1) and sewage
470 samples in Brazil [53] (SI Figure 3). Because bastroviruses have genomic features that
471 resemble hepeviruses [54], both these contigs had matches to invertebrate and vertebrate-
472 associated hepeviruses such that their true hosts could not be easily determined.

473 We also examined whether there were any differences in alpha and beta diversity between
474 native and invasive freshwater fishes. Accordingly, we found no association between host
475 origin (i.e. invasive or native) and virome abundance ($p = 0.390$). When assessing alpha
476 diversity, we similarly found no association between host origin and virome richness ($p =$
477 0.626) nor Shannon diversity ($p = 0.425$). Likewise, this result was also observed when
478 examining beta diversity ($p = 0.602$).

479 **Associations between host ecology and non-vertebrate viruses**

480 To assess any associations between host ecology and non-vertebrate viruses, we similarly
481 performed GLMs using the aforementioned ecological factors as a negative control. As
482 expected, this revealed no association between non-vertebrate viral abundance and host
483 taxonomy ($p = 0.200$), host geography ($p = 0.101$), host origin ($p = 0.998$), water turbidity (p
484 $= 0.421$), water temperature (0.282) and water pH ($p = 0.343$). We similarly found no
485 association between non-vertebrate virome richness and host taxonomy ($p = 0.204$), host
486 geography ($p = 0.090$), host origin ($p = 0.675$), water turbidity ($p = 0.398$), water temperature
487 ($p = 0.072$) and water pH ($p = 0.461$). We also found no evidence for an association
488 between Shannon diversity and host taxonomy ($p = 0.691$), host geography ($p = 0.173$),
489 host origin ($p = 0.876$), water turbidity ($p = 0.571$), water temperature ($p = 0.334$) and water
490 pH ($p = 0.578$). This was also observed when assessing statistical associations between
491 beta diversity and host species ($p = 0.684$), host origin ($p = 0.239$) and host geography ($p =$
492 0.501).

493

494

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

495 # Discussion

496 Our meta-transcriptomic viral survey of native and invasive fish across the Murray-Darling
497 Basin, Australia, revealed a high diversity and abundance of viruses, including the
498 identification of 45 novel virus species that infected seemingly healthy fish or non-
499 vertebrate hosts in the freshwater environment. Crucially, however, we observed no clear
500 examples of recent cross-species transmission among fish hosts, particularly between
501 invasive and native species, nor any evidence for the presence of *CyHV-3* from a total of 36
502 RNA sequencing libraries. Hence, these data provide further evidence of the absence of
503 *CyHV-3* in Australia [7]. Similarly, our analysis failed to detect other cypriniviruses (i.e.
504 *CyHV-1*, *CyHV-2*), despite previous reports of the presence of *CyHV-2* in the Murray-
505 Darling Basin [75, 76].

506 The only instance of co-occurrence of viruses from the same family in both invasive and
507 native species were the presence of arenaviruses in native carp-gudgeon and invasive
508 mosquitofish. However, these viruses were so divergent that they likely represent ancient
509 common ancestry rather than recent cross-species transmission (Figure 4). Eastern
510 mosquitofish, introduced into Australia during the early 1920s to control mosquito
511 populations [61], are now widespread across the Murray-Darling Basin and have become a
512 successful invasive species [62]. Their abundance primarily impacts smaller native fish
513 (such as carp-gudgeon, rainbowfish and hardyheads) since they typically outcompete these
514 species and disrupt food webs [62]. As well as being highly divergent, *western carp-*
515 *gudgeon arenavirus* formed a basal clade with a recently discovered arenavirus in the
516 pygmy goby sampled from an Australian coral reef [37], a member of the same fish order
517 (*Gobiiformes*). These data further suggest that arenaviruses may have been circulating in
518 Gobiiform fishes (gobies and gudgeons) in Australia prior to the introduction of eastern
519 mosquitofish.

520 We also identified a novel coronavirus - *Murray-Darling carp letovirus* – that shared 50.7%
521 amino acid sequence similarity with *Pacific salmon nidovirus* and 46.2% amino acid
522 similarity with gammacoronaviruses. The *Coronaviridae* (order *Nidovirales*) can be split into
523 two subfamilies: the *Orthocoronavirinae*, associated with birds and mammals and the
524 *Letovirinae* associated with amphibians [57]. That both *Murray-Darling carp letovirus* and
525 *Pacific salmon nidovirus* formed a sister clade to only member of the *Letovirinae* subfamily,
526 *Microhyla letovirus* (genus *Alphaletovirus*) identified in the ornamental pygmy frog
527 (*Microphyla fissipes*) (Figure 6) [57], suggests that fish may be common and ancient hosts
528 for the *Letovirinae*. It is also notable that *Murray-Darling carp letovirus* and *Pacific salmon*

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

529 *nidovirus* are highly divergent from the other *Nidovirales* that are known to infect fish (e.g.
530 *Chinook salmon bafinivirus*) [56].

531 The phylogenetic range of the *Chuviridae* largely incorporates invertebrate hosts with
532 diverse genomes (segmented, unsegmented and circular) [64]. Recently, chuviruses have
533 been discovered in vertebrates, all possessing three segments [18, 41]. The novel chuvirus
534 detected here in the unspecked hardyhead displayed these genomic features with the L
535 gene (RdRp), S gene (glycoprotein) and N gene (nucleoprotein) all related to fish and reptile
536 viruses (Figure 5) [18, 41]. The phylogenetic position of this vertebrate clade suggests the
537 ancestors of the viruses may be of invertebrate origin, particularly those that inhabit aquatic
538 ecosystems. For instance, the closest related invertebrate viruses were *Wenzhou crab virus*
539 [64], *Imjin River virus* (mosquitos) [65] and *Atrato chu-like virus* (mosquitos). Similarly,
540 chuvirus endogenous viral elements have been detected in several freshwater fish species
541 [18].

542 We identified a novel flavivirus in western carp-gudgeon across the Bogan River. This virus
543 falls basal to mammalian vector-borne viruses in phylogenetic trees, grouping with viruses
544 from other vertebrate hosts including *Cyclopterus lumpus virus, Tamana bat virus and*
545 *Wenzhou shark flavivirus*. Although *western carp-gudgeon flavivirus* was detected in
546 apparently healthy fish, *in vivo* flavivirus replication was recently demonstrated in Chinook
547 salmon (*Oncorhynchus tshawytscha*) that were associated with abnormal mortalities in the
548 Eel River, California [40]. While there is still no clear link between flavivirus infection,
549 transmission and disease in aquatic hosts, these data suggest that flaviviruses may be
550 common in fish species. Moreover, the basal phylogenetic positions of aquatic flaviviruses
551 also suggests that these viruses may be the ancestors of notable vector-borne viruses
552 (Figure 7). Nevertheless, gaps still remain in the evolutionary history of the genus *Flavivirus*
553 and will likely be bridged with additional metagenomic studies.

554 In broad terms, the evolutionary histories of many vertebrate viral families appear to
555 generally follow patterns of long-term virus-host co-divergence, albeit with regular cross-
556 species transmission [18,19]. This evolutionary pattern can be observed in the phylogenies
557 of the cultervirus, poxvirus and arenaviruses identified here. The *Bornaviridae* contain three
558 genera with 11 currently classified viral species that infect mammals, birds and reptiles [67].
559 The only fish virus identified to date falls within the genus *Cultervirus*, comprising
560 *Sharpbelly cultervirus* from China [18]. We identified this virus (i.e. transcripts with 93% L
561 gene amino acid similarly) in common carp in Australia. Intriguingly, both fish hosts are
562 members of the *Cyprinidae* that date as far back as the Cretaceous to Jurassic periods [68,

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

563 69]. Recent molecular clock dating using endogenous viral elements also showed that
564 culterviruses likely emerged early on during the course of vertebrate evolution, more than
565 50 million years ago [18].

566 Patterns of long-term virus-host co-divergence can also be seen in the evolutionary history
567 of the *Chordopoxvirinae*. *Western carp-gudgeon poxvirus* expands the host range of the
568 *Chordopoxvirinae* subfamily within the *Poxviridae*, forming a highly divergent clade with the
569 only other fish-infecting chordopoxvirus discovered to date, *salmon gill poxvirus (SGPV)*
570 (Figure 9)*.* Since its classification in 2015, several cases of *SGPV* have been identified in
571 farmed salmon with complex gill disease, although the reservoir host is unknown [55]. The
572 phylogeny of the *Chordopoxvirinae* mirrors that of vertebrate hosts, strongly suggesting
573 long-term virus-host co-divergence (Figure 9). Similarly, the phylogeny of the *Arenaviridae*
574 displays a basal fish clade that is characterised by long branches with a large degree of
575 divergence (Figure 4).

576 On this evolutionary backbone of ancient virus-host co-divergence, we also detected cases
577 of cross-species virus transmission during evolutionary history, although the time-scales of
578 these events are uncertain. For example, we discovered a novel reovirus that formed a
579 basal divergent clade to other fish viruses within the genus *Aquareovirus* (Figure 8). *Murray-*
580 *Darling carp reovirus* was more closely related to viruses that infect scaldfish [18] rather
581 than other cyprinid hosts, which are highly susceptible to reovirus infection (e.g. 80%
582 mortality in grass carp) [42]. These patterns were also observed in the phylogeny of the
583 *Rhabdoviridae*, with *Murray-Darling carp rhabdovirus* forming a distinct phylogenetic clade
584 with other recently discovered rhabdoviruses in fish, the big brown bat (NCBI/Genbank:
585 QPO14166.1) and the spotted paddle-tail newt [18]. Rhabdoviruses exhibit a very broad
586 host range including invertebrates, plants, mammals, fish, amphibians, birds and reptiles
587 [70]. Notable among the *Rhabdoviridae* are the lyssaviruses that can cause high mortality in
588 human populations (e.g. *rabies lyssavirus*). Intriguingly, *Murray-Darling carp rhabdovirus* and
589 its closest relatives form a sister clade to the genus *Lyssavirus*, suggesting these viruses
590 may have a fish-infecting ancestor (Figure 4).

591 Although carp are widespread and abundant across the Murray-Darling Basin, they
592 displayed lower viral abundance than some of the other hosts sampled (Figure 2). This
593 could be partly explained by the large phylogenetic distance between carp and other fish in
594 the Murray-Darling Basin. For instance, aside from bony herring, all the fish studied here are
595 members of the *Acanthopterygii* (Percomorpha). This could also explain why invasive
596 mosquitofish harboured similar viruses to native gudgeon species (e.g. arenaviruses).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

597 Although not always the case [22], cross-species virus transmission often occurs between
598 phylogenetically related hosts, particularly those that display conserved cell receptors [66,
599 71]. In addition, it has been widely suggested that introduced populations are associated
600 with a lower pathogen prevalence and diversity than native species [73, 74]. For example,
601 because invasive species are often established from a small founder population, they likely
602 acquire only a small proportion of pathogens in the novel environment [73, 74]. Once a
603 species rapidly becomes invasive, the diversity of pathogens in this population should
604 remain small, such that the lack of disease likely facilitates the success of invasive species
605 [73, 74].

606 It is important to note, however, that there were necessary variations within our sampling.
607 For instance, carp and native fish species were sampled together at 10 out of 13 sites, with
608 carp sampled from all 13 sites (Figure 1). In addition, all other fish species were sampled
609 from 1-5 sites. While an artifact of the distribution of the fishes, such gaps limit the power of
610 our statistical analyses and perhaps prevent the detection of ecological associations on
611 virome composition within host species, including between invasive and native fish. In
612 addition, due to animal ethics constraints, we were limited to only a subset of native fish
613 species. Nevertheless, the native species examined in this study are generally those
614 present in the highest densities.

615 Finally, our analysis detected no viruses that are listed as reportable notifiable aquatic
616 diseases in the Murray-Darling Basin [72]. Such notifiable aquatic diseases include
617 epizootic haematopoietic necrosis virus (EHNV - *Iridoviridae*) and Spring viraemia of carp
618 (SVC – *Rhabdoviridae*). EHNV is known to cause high-impact infections in redfin perch and
619 is capable of infecting other freshwater fish in the Murray-Darling Basin, including silver
620 perch (*Bidyanus bidyanus*), Macquarie perch (*Macquaria australasica*), Murray-Darling
621 rainbowfish, freshwater catfish (*Tandanus tandanus*) and invasive mosquitofish [47].
622 Although thought to be endemic to the Murray-Darling Basin (upper Murrumbidgee River),
623 EHNV was last reported in 2012 [63]. Similarly, we did not detect the emerging dwarf
624 gourami iridovirus (*Iridoviridae*) that causes infectious spleen and kidney necrosis in several
625 species of native Australian fish [79, 80].

626 In sum, our metagenomic surveillance revealed a marked lack of virus exchange between
627 native and invasive fish species in the Murray-Darling Basin, including those viruses found
628 in invasive common carp. At face value these data suggest that there is minimal virus
629 transmission from common carp to native fish species, although more extensive sampling
630 is needed to fully address this issue. By investigating the viromes of native and invasive

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

631 fishes, we provide the first data on viruses that naturally circulate in a 2,200 km river
632 system, enhancing our understanding of the evolutionary history of vertebrate viruses.

633

634 Data Availability

635 All sequence reads generated in this project are available under the NCBI Short Read
636 Archive (SRA) under BioProject PRJNA701716 and all consensus virus genetic
637 sequences have been deposited in GenBank under accessions MW645018-MW645046.

638

639 Acknowledgements

640 This work was funded by a Recreational Fishing Trust grant (SF024) awarded to JLG, JEW
641 and ECH; an ARC Discovery grant (DP200102351) awarded to ECH and JLG; a Macquarie
642 University grant awarded to JLG; and the Department of Biological Sciences at Macquarie
643 University. ECH is funded by ARC Australian Laureate Fellowship (FL170100022). J.L.G. is
644 funded by a New Zealand Royal Society Rutherford Discovery Fellowship (RDF-20-UOO-
645 007). We thank Justin Stanger, Patrick Martin, Karen Scott, Michael Rodgers, Duncan
646 McLay and Nick O'Brien from the Department of Primary Industries for their help with fish
647 sampling. We also thank efishalbum.com for providing us with the fish images used in
648 Figures 1 and 3, which were used with permission.

649

650 References

651 1. World Wildlife Fund (WWF). The world's forgotten fishes. 2021 Feb 23. Available from:
652 https://wwf.panda.org/discover/our_focus/freshwater_practice/the_world_s_forgotten_fishe
653 s/

654 2. Barrett J, Bamford H, Jackson P. Management of alien fishes in the Murray-Darling
655 Basin. Ecol Manag Restor. 2014;15(1): 51-56.

656 3. Koehn JD. Carp (Cyprinus carpio) as a powerful invader in Australian waterways. Freshw
657 Biol. 2004;49(7): 882-894.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

658 4. Forsyth D, Koehn J, MacKenzie D, Stuart I. Population dynamics of invading freshwater
659 fish: common carp (Cyprinus carpio) in the Murray-Darling Basin, Australia. Biol Invasions.
660 2013;15(2): 341-354.

661 5. Vilizzi L, Thwaites LA, Smith BB, Nicol JM, Madden CP. Ecological effects of common
662 carp (Cyprinus carpio) in a semi-arid floodplain wetland. Mar Freshw Res. 2014;65(9): 802-
663 817.

664 6. Hayes KR, Leung B, Thresher R, Dambacher JM, Hosack GR. Meeting the challenge of
665 quantitative risk assessment for genetic control techniques: a framework and some
666 methods applied to the common Carp (Cyprinus carpio) in Australia. Biol Invasions.
667 2014;16(6): 1273-1288.

668 7. McColl KA, Sunarto A, Slater J, Bell K, Asmus M, Fulton W, et al. Cyprinid herpesvirus 3
669 as a potential biological control agent for carp (Cyprinus carpio) in Australia: susceptibility
670 of non-target species. J Fish Dis. 2017;40(9): 1141-1153.

671 8. McColl KA, Cooke BD, Sunarto A. Viral biocontrol of invasive vertebrates: lessons from
672 the past applied to cyprinid herpesvirus-3 and carp (Cyprinus carpio) control in Australia.
673 Biol Control. 2014;72: 109-117.

674 9. Matsui K, Honjo M, Kohmatsu Y, Uchii K, Yonekura R, Kawabata ZI. Detection and
675 significance of koi herpesvirus (KHV) in freshwater environments. Freshw Biol. 2008;53(6):
676 1262-1272.

677 10. Michel B, Fournier G, Lieffrig F, Costes B, Vanderplasschen A. Cyprinid herpesvirus 3.
678 Emerg Infect Dis. 2010;16(12): 1835-1843.

679 11. Shimizu T, Yoshida N, Kasai H, Yoshimizu M. Survival of koi herpesvirus (KHV) in
680 environmental water. Fish Pathol. 2006(4): 153-157.

681 12. Bergmann SM, Sadowski J, Kiełpiński M, Bartłomiejczyk M, Fichtner D, Riebe R, et al.
682 Susceptibility of koi × crucian carp and koi × goldfish hybrids to koi herpesvirus (KHV) and
683 the development of KHV disease (KHVD). J Fish Dis. 2010;33(3): 267-272.

684 13. Kopf RK, Boutier M, Finlayson CM, Hodges K, Humphries P, King A, et al. Biocontrol in
685 Australia: can a carp herpesvirus (CyHV-3) deliver safe and effective ecological restoration?
686 Biol Invasions. 2019;21(6): 1857-1870.

687 14. Iouze M, Davidovich M, Diamant A, Kotler M, Dishon A. The outbreak of carp disease
688 caused by CyHV-2 as a model for new emerging viral diseases in aquaculture: a review.
689 Ecol Res. 2011;26: 885-892

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

690 15. Tolo IE, Padhi SK, Williams K, Singh V, Halvorson S, Sunil KM. Susceptibility of
691 Pimephales promelas and Carassius auratus to a strain of koi herpesvirus isolated from wild
692 Cyprinus carpio in North America. Sci Rep. 2021;11: 1985.

693 16. Shi M, Zhang Y-Z, Holmes EC. Meta-transcriptomics and the evolutionary biology of
694 RNA viruses. Virus Res. 2018;243: 83-90.

695 17. Zhang Y-Z, Shi M, Holmes EC. Using metagenomics to characterize an expanding
696 virosphere. Cell. 2018;172(6): 1168-1172.

697 18. Shi M, Lin X-D, Chen X, Tian J-H, Chen L-J, Li K, et al. The evolutionary history of
698 vertebrate RNA viruses. Nature. 2018;561: 197-202.

699 19. Zhang Y-Z, Wu W-C, Shi M, Holmes EC. The diversity, evolution and origins of
700 vertebrate RNA viruses. Curr Opin Virol. 2018;31: 9-16.

701 20. Dudgeon D, Arthington AH, Gessner MO, Kawabata ZI, Knowler DJ, Lévêque C, et al.
702 Freshwater biodiversity: importance, threats, status and conservation challenges. Biol Rev
703 Camb Philos Soc. 2006;81(2): 163-182.

704 21. Johnson PTJ, Paull SH. Ecology and emergence of diseases in fresh waters. Freshw
705 Biology. 2011;56(4): 638-657.

706 22. Parrish CR, Holmes EC, Morens DM, Park E-C, Burke DS, Calisher CH, et al. Cross-
707 species virus transmission and the emergence of new epidemic diseases. Microbiol Rev.
708 2008;72(3): 457-470.

709 23. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence
710 data. Bioinformatics. 2014;30(15): 2114-2120.

711 24. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood P, Bowden J, et al. De novo
712 transcript sequence reconstruction from RNA-seq using the Trinity platform for reference
713 generation and analysis. Nat Protoc. 2013;8(8): 1494.

714 25. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND.
715 Nat Methods. 2014;12(1): 59-60.

716 26. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious
717 basic: an integrated and extendable desktop software platform for the organization and
718 analysis of sequence data. Bioinformatics. 2012;28(12): 1647-1649.

719 27. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7:
720 improvements in performance and usability. Mol Biol Evol. 2013;30(4): 772-780.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

721 28. Lagkouvardos I, Fischer S, Kumar N, Clavel T. Rhea: a transparent and modular R
722 pipeline for microbial profiling based on 16S rRNA gene amplicons. PeerJ. 2017;5(1):
723 e2836.

724 29. Hothorn T, Bretz F, Westfall P. Simultaneous inference in general parametric models.
725 Biom J. 2008;50(3): 346-363.

726 30. McMurdie PJ, Holmes S. phyloseq: An R Package for reproducible interactive analysis
727 and graphics of microbiome census data. PLoS One. 2013;8(4): e61217.

728 31. Dixon P. VEGAN, a package of R functions for community ecology. J Veg Sci.
729 2003;14(6): 927-930.

730 32. Valero-Mora PM. ggplot2: Elegant graphics for data analysis. J Stat Softw. 2010;35.

731 33. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated
732 alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15): 1972-
733 1973.

734 34. Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective
735 stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol.
736 2015;32(1): 268-274.

737 35. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder:
738 fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6): 587-
739 589.

740 36. Abudurexiti A, Adkins S, Alioto D, Alkhovsky S, Avšič-Županc T, Ballinger M, et al.
741 Taxonomy of the order Bunyavirales: update 2019. Arch Virol. 2019;164(7): 1949-1965.

742 37. Geoghegan JL, Di Giallonardo F, Wille M, Ortiz-Baez AS, Costa VA, Ghaly T, et al.
743 Virome composition in marine fish revealed by meta-transcriptomics. Virus Evol. 2021;7(1):
744 veab005.

745 38. Skoge R, Brattespe J, Økland A, Plarre H, Nylund A. New virus of the family Flaviviridae
746 detected in lumpfish (Cyclopterus lumpus). Arch Virol. 2018;163(3): 679-685.

747 39. de Lamballerie X, Crochu S, Billoir F, Neyts J, de Micco P, Holmes EC, et al. Genome
748 sequence analysis of Tamana bat virus and its relationship with the genus Flavivirus. J Gen
749 Virol. 2002;83(10): 2443-2454.

750 40. Soto E, Camus A, Yun S, Kurobe T, Leary JH, Rosser TG, et al. First isolation of a novel
751 aquatic flavivirus from chinook salmon (Oncorhynchus tshawytscha) and its replication in a
752 piscine animal model. J Virol. 2020;94(15): e00337-20.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

753 41. Argenta FF, Hepojoki J, Smura T, Szirovicza L, Hammerschmitt ME, Driemeier D, et al.
754 Identification of reptarenaviruses, hartmaniviruses, and a novel chuvirus in captive native
755 Brazilian boa constrictors with boid Inclusion body disease. J Virol. 2020;94(11): e00001-
756 20.

757 42. Chen G, He L, Luo L, Huang R, Liao L, Li Y, et al. Transcriptomics sequencing provides
758 insights into understanding the mechanism of grass carp reovirus infection. Int J Mol Sci.
759 2018;19(2): 488.

760 43. Tordo N, Poch O, Ermine A, Keith G, Rougeon F. Completion of the rabies virus
761 genome sequence determination: highly conserved domains among the L (polymerase)
762 proteins of unsegmented negative-strand RNA viruses. Virology. 1988(2): 565-576.

763 44. Geoghegan JL, Di Giallonardo F, Cousins K, Shi M, Williamson JE, Holmes EC. Hidden
764 diversity and evolution of viruses in market fish. Virus Evol. 2018;4(2): vey031.

765 45. Shi M, Lin X-D, Tian J-H, Chen L-J, Chen X, Li C-X, et al. Redefining the invertebrate
766 RNA virosphere. Nature. 2016;540(7634): 539-543.

767 46. Hameed AS, Ninawe AS, Nakai T, Chi SC, Johnson KL. ICTV Virus Taxonomy Profile:
768 Nodaviridae. J Gen Virol. 2019;100(1): 3-4.

769 47. Becker JA, Tweedie A, Gilligan D, Asmus M, Whittington RJ. Experimental infection of
770 Australian freshwater fish with epizootic haematopoietic necrosis virus (EHNV). J Aquat
771 Anim Health. 2013;25(1): 66-76.

772 48. Fichtner D, Philipps A, Groth M, Schmidt-Posthaus H, Granzow H, Dauber M, et al.
773 Characterization of a novel picornavirus isolate from a diseased European eel (Anguilla
774 anguilla). J Virol. 2013;87(19): 10895-10899.

775 49. Vinjé J, Estes MK, Esteves P, Green KY, Katayama K, Knowles NJ, et al. ICTV Virus
776 Taxonomy Profile: Caliciviridae. J Gen Virol. 2019;100(11): 1469-1470.

777 50. Mikalsen AB, Nilsen P, Frøystad-Saugen M, Lindmo K, Eliassen TM, Rode M, et al.
778 Characterization of a novel calicivirus causing systemic infection in Atlantic salmon (Salmo
779 salar L.): proposal for a new genus of Caliciviridae. PLoS One. 2014;9(9): e107132.

780 51. Gjessing MC, Yutin N, Tengs T, Senkevich T, Koonin E, Rønning HP, et al. Salmon gill
781 poxvirus, the deepest representative of the Chordopoxvirinae. J Virol. 2015;89(18): 9348-
782 9367.

783 52. Sadeghi M, Altan E, Deng X, Barker CM, Fang Y, Coffey LL, et al. Virome of >12
784 thousand Culex mosquitoes from throughout California. Virology. 2018;523: 74-88.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

785 53. Dos Anjos K, Nagata T, Melo FLd. Complete genome sequence of a novel bastrovirus
786 isolated from raw sewage. Genome Announc. 2017;5(40).

787 54. Reuter G, Boros Á, Tóth Z, Kapusinszky B, Delwart E, Pankovics P. Detection of a novel
788 RNA virus with hepatitis E virus-like non-structural genome organization in amphibian, agile
789 frog (Rana dalmatina) tadpoles. Infect Genet Evol. 2018;65: 112-116.

790 55. Gjessing MC, Thoen E, Tengs T, Skotheim SA, Dale OB. Salmon gill poxvirus, a recently
791 characterized infectious agent of multifactorial gill disease in freshwater and seawater-
792 reared Atlantic salmon. J Fish Dis. 2017;40(10): 1253-1265.

793 56. Cano I, Stone D, Savage J, Wood G, Mulhearn B, Gray J, et al. Isolation of a chinook
794 salmon bafinivirus (CSBV) in imported goldfish L. in the United Kingdom and evaluation of
795 its virulence in resident fish species. Viruses. 2020;12(5): 578.

796 57. Bukhari K, Mulley G, Gulyaeva AA, Zhao L, Shu G, Jiang J, et al. Description and initial
797 characterization of metatranscriptomic nidovirus-like genomes from the proposed new
798 family Abyssoviridae, and from a sister group to the Coronavirinae, the proposed genus
799 Alphaletovirus. Virology. 2018;524: 160-171.

800 58. Woo PCY, Lau SKP, Lam CSF, Tsang AKL, Hui S-W, Fan RYY, et al. Discovery of a
801 novel bottlenose dolphin coronavirus reveals a distinct species of marine mammal
802 coronavirus in Gammacoronavirus. J Virol. 2014;88(2): 1318-1331.

803 59. Mihindukulasuriya KA, Wu G, St. Leger J, Nordhausen RW, Wang D. Identification of a
804 novel coronavirus from a beluga whale by using a panviral microarray. J Virol. 2008;82(10):
805 5084-5088.

806 60. Mordecai GJ, Miller KM, Di Cicco E, Schulze AD, Kaukinen KH, Ming TJ, et al.
807 Endangered wild salmon infected by newly discovered viruses. eLife. 2019;8: e47615.

808 61. Ayres R, Pettigrove V, Hoffmann A. Low diversity and high levels of population genetic
809 structuring in introduced eastern mosquitofish (Gambusia holbrooki) in the greater
810 Melbourne area, Australia. Biol Invasions. 2010;12(11): 3727-3744.

811 62. Macdonald JI, Tonkin ZD, Ramsey DSL, Kaus AK, King AK, Crook DA. Do invasive
812 eastern gambusia shape wetland fish assemblage structure in south-eastern Australia? Mar
813 Freshw Res. 2012;63(8): 659-671.

814 63. Becker JA, Gilligan D, Asmus M, Tweedie A, Whittington RJ. Geographic distribution of
815 (EHNV) in freshwater fish in south eastern Australia: lost opportunity for a notifiable
816 pathogen to expand its geographic range. Viruses. 2019;11(4): 315.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

817 64. Jun-Hua T, Xian-Dan L, Yan-Jun K, Liang-Jun C, Xin-Cheng Q, Xu J, et al.
818 Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of
819 negative-sense RNA viruses. eLife. 2015;4(4): e05378.

820 65. Hang J, Klein TA, Kim H-C, Yang Y, Jima DD, Richardson JH, et al. Genome sequences
821 of five arboviruses in field-captured mosquitoes in a unique rural environment of South
822 Korea. Genome Announc. 2016;4(1).

823 66. Longdon B, Brockhurst MA, Russell CA, Welch JJ, Jiggins FM. The evolution and
824 genetics of virus host shifts. PLoS Pathog. 2014;10(11): e1004395

825 67. Amarasinghe G, Ayllón M, Bào Y, Basler C, Bavari S, Blasdell K, et al. Taxonomy of the
826 order Mononegavirales: update 2019. Arch Virol. 2019;164(7): 1967-1980.

827 68. Hughes LC, Ortí G, Huang Y, Sun Y, Baldwin CC, Thompson AW, et al. Comprehensive
828 phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data.
829 Proc Natl Acad Sci. 2018;115(24): 6249-6254.

830 69. Betancur-R R, Wiley EO, Arratia G, Acero A, Bailly N, Miya M, et al. Phylogenetic
831 classification of bony fishes. BMC Evol Biol. 2017;17(1).

832 70. Bourhy H, Cowley JA, Larrous F, Holmes EC, Walker PJ. Phylogenetic relationships
833 among rhabdoviruses inferred using the L polymerase gene. J Gen Virol. 2005;86(10): 2849-
834 2858.

835 71. Dennehy JJ. Evolutionary ecology of virus emergence. Ann N Y Acad Sci. 2017;1389(1):
836 124-146.

837 72. Australian Health Committee (AHC). Australia's national list of reportable diseases of
838 aquatic animals. 2020 Sep 4. Available from:
839 https://www.agriculture.gov.au/animal/aquatic/reporting/reportable-diseases#finfish

840 73. Phillips BL, Kelehear C, Pizzatto L, Brown GP, Barton D, Shine R. Parasites and
841 pathogens lag behind their host during periods of host range advance. Ecol. 2010;91(3):
842 872-881.

843 74. Mark ET, Kevin DL, Andrew PD, Valerie JM, Armand MK. Introduced species and their
844 missing parasites. Nature. 2003;421(6923): 628-630.

845 75. Stephens F, Raidal, Sr., Jones B. Haematopoietic necrosis in a goldfish (Carassius
846 auratus) associated with an agent morphologically similar to herpesvirus. Aust Vet J.
847 2004;82(3): 167-169.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

848 76. Becker JA, Tweedie A, Rimmer A, Landos M, Lintermans M, Whittington RJ. Incursions
849 of Cyprinid herpesvirus 2 in goldfish populations in Australia despite quarantine practices.
850 Aquac. 2014;432: 53-59.

851 77. Lauber C, Seitz S, Mattei S, Suh A, Beck J, Herstein J, et al. Deciphering the origin and
852 evolution of hepatitis b viruses by means of a family of non-enveloped fish viruses. Cell
853 Host Microbe. 2017;22(3): 387-399.

854 78. Dill JA, Camus AC, Leary JH, Di Giallonardo F, Holmes EC, Ng TFF. Distinct viral
855 lineages from fish and amphibians reveal the complex evolutionary history of
856 Hepadnaviruses. J Virol. 2016;90(17): 7920-7933.

857 79. Go J, Lancaster M, Deece K, Dhungyel O, Whittington R. The molecular epidemiology
858 of iridovirus in Murray cod (Maccullochella peelii peelii) and dwarf gourami (Colisa lalia) from
859 distant biogeographical regions suggests a link between trade in ornamental fish and
860 emerging iridoviral diseases. Mol Cell Probes. Jun-Aug 2006;(3-4): 212-222.

861 80. Rimmer AE, Whittington RJ, Tweedie A, Becker JA. Susceptibility of a number of
862 Australian freshwater fishes to dwarf gourami iridovirus (Infectious spleen and kidney
863 necrosis virus). J Fish Dis. 2017;40(3): 293-310.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.25.432824; this version posted February 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

864 # Supplementary information

865 **SI Table 1.** Contig length and amino acid similarity of potentially novel vertebrate-
866 associated viruses identified in this study.

867 **SI Table 2.** RNA sequencing library details, including sampling site, replicates and RNA
868 concentration.

869 **SI Figure 1.** Phylogenetic relationships of the likely non-vertebrate viruses within the
870 *Rhabdoviridae, Picornaviridae, Tombusviridae* and *Narnaviridae*. Viruses identified here are
871 highlighted in blue. Maximum likelihood trees were estimated using amino acid sequences
872 of the RdRp gene. Trees were mid-point rooted for clarity only and bootstrap values are
873 represented as a percentage with branches scaled to amino acid substitutions per site. Tip
874 labels represent virus name and NCBI/GenBank accession IDs.

875 **SI Figure 2.** Phylogenetic relationships of the likely non-vertebrate viruses within the
876 *Nodaviridae, Phenuiviridae, Dicistroviridae* and *Permutotetraviridae*. Viruses identified here
877 are highlighted in blue. Maximum likelihood trees were estimated using amino acid
878 sequences of the RdRp gene. Trees were mid-point rooted for clarity only and bootstrap
879 values are represented as a percentage with branches scaled to amino acid substitutions
880 per site. Tip labels represent virus name and NCBI/GenBank accession IDs.

881 **SI Figure 3.** Phylogenetic relationships of bastrovirus sequences. Novel viruses are
882 represented as black circles. Maximum likelihood tree was estimated using amino acid
883 sequences of the ORF1 polypeptide. Trees were mid-point rooted for clarity only and
884 bootstrap values are represented as a percentage with branches scaled to amino acid
885 substitutions per site. Tip labels represent virus name and NCBI/GenBank accession.